UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/528,636	03/22/2005	Lionel Oisel	PF020128	7137

24498 7590 01/26/2009

Robert D. Shedd
Thomson Licensing LLC
PO Box 5312
PRINCETON, NJ 08543-5312

EXAMINER	
BLACK, LINH	

ART UNIT	PAPER NUMBER
2169	

MAIL DATE	DELIVERY MODE
01/26/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 10/528,636
Filing Date: March 22, 2005
Appellant(s): OISEL ET AL.

Joseph J. Opalach
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 11/17/2008 appealing from the Office action mailed 4/18/2008.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

The amendment after final rejection filed on 6/18/08

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is correct.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

5,821,945	Yeo et al.	10-1998
7,054,367	Oguz et al.	5-2006
7,212,201	Geiger et al.	5-2007

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-2 and 3-6 are rejected under 35 U.S.C. 103(a) as being unpatentable over Yeo et al. (US 5821945), in view of Oguz et al. (US 7054367), and further in view of Geiger et al. (US 7212201).

As per claim 1, Yeo et al. teach:

Method of clustering images of a video sequence consisting of shots and represented by a graph-like structure – fig. 1; col. 4, last paragraph to col. 5, line 2.

a node of the graph representing a shot or a class of shots defined by key images and the nodes being connected by edges – col. 4, 1st paragraph (each node represents a cluster of shots, which are considered a scene in the general sense. A directed edge is drawn from node U to node W if there is a shot represented by node U that immediately precedes some shot represented by node W).

comprising the following iteration: selecting of an edge ak connecting nodes ni and nj – col. 4, lines 23-29; col. 5, lines 32-52 (the nodes capture the core contents of the video while the edges capture its structure. The browsing approach thus is based on both content and structure of a complex video selection); col. 9, last paragraph (iteration).

calculating of the potential of node nm, merging of the two nodes ni and nj, the attributes of the key images defining the class of shots of node ni and those of the key images defining the class of shots of node nj– col. 2, lines 39-55 (long sequences of related shots can be telescoped into a small number of key frames which represent the repeatedly appearing shots in the scene); col. 9, lines 3-18 (the present system algorithm first groups the pair of shots by their proximity values; "proximity value" can be interpreted as equivalent to the "potential"); col. 6, lines 14-32 (clustering of shots is equivalent to merging of nodes); col. 8, line 43 to col. 9, line 18; col. 5, lines 32-57 (the shots that exhibit visual, spatial and temporal similarities are then clustered into scenes…the primitive attributes of the shots contribute the major clustering criteria at the initial stage of the scene). Yeo et al. teach grouping shots by their proximity values and it is preferred to have a shot left as a single cluster/new cluster than to have it grouped into other clusters not in close match – col. 9, 1st paragraph. Yeo et al. do not

suggest distance or temporal distance of the key images/frames. Oguz et al. teach attributes are compared to at least one length threshold to detect a scene change in the MPEG video sequence – col. 6, 2nd paragraph; compute a degree of coincidence between significant edges in a current frame and significant edges in a prior frame to within a distance and temporal distance ...- col. 8, 2nd paragraph. Thus, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Yeo's teaching and Oguz's teaching to allow different categorization/clustering method to be utilized. However, Yeo and Oguz do not suggest the sum of the potentials of the nodes and of the edges, is less than an energy of the graph before merging. Geiger et al. teach the result of merging two nodes in the graph shown in fig. 7b; the way that the edge weights are defined, the minimum cut corresponds to the optimal segmentation that is it has the minimum sum of equation – col. 6, lines 29-36; the weight of an edge connecting some node x and a merged node y is given by the sum of weights of all edges that connect node x and all nodes that are merged into y – col. 20, lines 23-39. Thus, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Yeo's teaching, Oguz's teaching, and Geiger's teaching in order to improve efficiency when nodes that are likely never be separated, thus, merge the two nodes into a cluster for better manipulation of data.

As per claim 2, Yeo et al. teach:

wherein the graph is initialized by assigning a node to each shot and in that edges are created from one node to another node if the shots relating to these nodes are

separated by a predetermined maximum number T of shots – col. 4, 1st paragraph; col. 2, lines 3-23; col. 8, lines 27-41 (It is important to balance the two goals: to preserve as much of the temporal variations as possible and to reduce the computing load needed to process many video frames in a given shot. In the present system, the inventors chose a good but nevertheless greatly reduced representative set of frames to represent a video shot).

As per claim 3, Yeo et al. do not suggest temporal distance. Oguz et al. teach attributes are compared to at least one length threshold to detect a scene change in the MPEG video sequence – col. 6, 2nd paragraph; compute a degree of coincidence between significant edges in a current frame and significant edges in a prior frame to within a distance and temporal distance ...- col. 8, 2nd paragraph; to detect edges, ...code length is compared to a threshold length to produce a bit indicating the presence or absence of an edge... if the threshold length is too large, only the strongest edges will be detected. If the threshold length is too small, some features will mistakenly be detected as edges, the false alarm rate will increase...- col. 7, lines 18-39. Thus, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Yeo's teaching and Oguz's teaching to allow different categorization/clustering method to be utilized.

As per claim 6, Yeo et al. teach grouping shots by their proximity values and it is preferred to have a shot left as a single cluster/new cluster than to have it grouped into

other clusters not in close match... automatic clustering schemes for scene transition graph building can be made at multiple levels. At each level, a different criterion is impose...In the top levels of the hierarchy, subgraph properties and temporal structures, such as discovering repeated self-loops and subgraph isomorphism, can be explored to further condense the graph – col. 9, lines 1-67. (the automatic/repeating clustering process would stop after the potential merging/clustering of two nodes give rise to an increase in energy...Yeo et al. teach "it is preferred to have a shot left as a single cluster than to have it grouped into other clusters not in close match" – col. 9, lines 16-18).

Allowable Subject Matter

Claims 4-5 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

Applicant's arguments filed 1/22/08 have been fully considered but they are not persuasive. Regarding the argument "nowhere does Oguz describe or suggest that the node potential is a function of this temporal distance...key images within the sequence." Examiner disagrees. The limitation of claim 1 teaches "...a function of distances between the attributes of the key images...". Oguz describes temporal distance as cited in the office action. Key frames show scenes or features/attributes of a video sequence

change - col. 5, line 39 to col. 6, line 35 wherein images represented by the I-frames – col. 3, lines 25-41.

Regarding the Applicants' argument on page 6, last paragraph, how the minimum function is calculated does not seem to be disclosed in the rejected claims. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

(10) Response to Argument

Argument A : claims 1-3 and 6

In response to the Appellant's argument on page 9 that "Examiner is wrong and Oguz does not describe temporal distance as claimed...there is no mention of a temporal distance relating to images as required...Applicants also note ...on col. 8, 2nd paragraph of Oguz. Again with respect to the requirements of Applicants' claim 1, the Examiner is wrong. This distance is only a spatial distance. Applicants do note that the distance disclosed on line 14 of Oguz is a temporal distance between the current and prior frames. But, this temporal distance is used to determine the amount of motion in the scene (OGUZ, line 15)..." Examiner disagrees.

In the preamble of claim 1, the Appellant discloses "Method of clustering images of a video sequence consisting of shots and represented by a graph-like structure, a node of the graph representing a shot or a class of shots defined by key images and the nodes being connected by edges".

Yeo et al. discloses "classify a long video sequence into story units, based on its content. Scene change detection (also called TEMPORAL SEGMENTATION of video) give sufficient indication of when a new shot starts and ends,..Beyond temporal segmentation of video, one known browser users Rframes (representative frames) to organize the visual contents of the video clips. Rframes may be grouped according to various criteria to aid the user in identifying the desired material. The user can select a key frame, and the system then uses various criteria to search for similar key frames and present them to the user as a group" - col. 1, last two paragraphs.

Yeo et al. also discloses "the story structure is modeled with a hierarchical scenes transition graph, and the scenic structure is extracted using visual and temporal

information with no priori knowledge of the content - the structure is discovered automatically. A hierarchical scene transition graph reflects the decomposition of the video into acts, scenes and shots. Such a hierarchical view of the video provides an effective means for browsing the video content, since long sequences of related shots can be telescoped into a small number of **key frames which represent the repeatedly appearing shots** in the scene" – col. 2, lines 35-45.

In col. 4, lines 23-50, Yeo et al. discloses "This is a hierarchical organization in time of the collection of shots. At the lowest level, each node Vo,i represents L shots; a directed edge connects Vo,i to Vo,i+1...Such a tree hierarchy permits a user to have a coarse-to-fine view of the entire video sequences...In this case, shots that are similar to each other are clustered together. Relations between clusters are governed by temporal ordering of shots within the two clusters..."; "From the clustering results and the temporal information associated with each shot, the system proceeds to build the graphs, with nodes representing scenes and edges representing the progress of the story from one scene to the next. The nodes capture the core contents of the video while the edges capture its structure" – col. 5, lines 36-41.

Therefore, Yeo et al. does disclose the clustering/classifying of shots in video sequences by using the shots' temporal information, and also the relations between clusters are governed by temporal ordering of shots within the clusters (cited cols. 4 and 5 above). Yeo also discloses "long sequences of related shots can be telescoped into a small number of **key frames which represent the repeatedly appearing shots** in the scene" – col. 2, lines 35-45. Because "key frames" represent "repeatedly appearing

shots", a "key frame" does represent a class of shots. "Key frames" thus, are equivalent to the Appellant's limitation "key images". The limitation "temporal ordering of shots" seems to equivalent to "the ordering of shots' temporal values/distances". Examiner combined Yeo's teachings with Oguz's teaching because Yeo et al. does not explicitly disclose the limitation "distance".

The Appellant noted on page 9, last paragraph that "Applicants do note that the distance disclosed on line 14 of Oguz is a temporal distance between the current and prior frames...This temporal distance allows one to determine the number of blocks to consider in order to calculate similarities...However, Yeo discloses "long sequences of related shots can be telescoped into a small number of **key frames which represent the repeatedly appearing shots** in the scene" – col. 2, lines 35-45. Because "key frames" represent "repeatedly appearing shots", a "key frame" does represent a class of shots. "Key frames" thus, are equivalent to the Appellant's limitation "key images".

Examiner combined Oguz's teaching with Yeo's teaching in order to show that using of temporal information/distance in clustering shots/frames in video sequences are not novel in the technological art. Oguz discloses in col. 8, lines 8-32 the detecting of scene change, the usage of TEMPORAL DISTANCE between the frames, the matching of edges between the frames etc...As Yeo discloses key frames represent the repeatedly appearing shots in the shots-clustering process – Yeo, col. 2, lines 35-45. Therefore, the usage of TEMPORAL DISTANCE to study scene change and to cluster shots are not novel in the technological art.

Regarding arguments on page 10, Appellant states in lines 12-13 that "In Applicants' claimed invention, the potential of the edge linking two key frames is a function of the temporal distance." Appellant seems use "key frames" and "key images" interchangeably here. Appellant also contents that "In Applicants' claimed invention temporal distance is not used to define criteria (such as vicinity)...The way the attribute differences are calculated is not specified in claim 1...". As cited in the Office action above, "Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993)". Therefore, the argument above is against a disclosure in the specification not the claim's limitation.

Appellant argues on page 11 regarding "In Yeo there is no description or suggestion about temporal distances and merging of nodes according to potentials which are function of these temporal distances as claimed". Examiner disagrees.

In addition to paragraphs cited above regarding Yeo's teaching of clustering shots into a hierarchical graph using temporal information/ordering of shots, Yeo discloses in col. 4, 1st paragraph the "grouping of shots at the lowest level of hierarchy. The collection f shots is partitioned into nodes of Go; each node represents a cluster of shots, which are considered a scene in the general sense. A directed edge is drawn from node U to W if there is a shot represented by node U that immediately precedes some shot represented by node W. Further grouping into other levels of the hierarchy is defined in a similar fashion in property. The edge relationship induced by TEMPORAL PRECEDENCE at level 0 is preserved as one moves up the hierarchy."

In col. 4, lines 23-50, Yeo et al. discloses "This is a hierarchical organization in time of the collection of shots. At the lowest level, each node Vo,i represents L shots; a directed edge connects Vo,i to Vo,i+1...Such a tree hierarchy permits a user to have a coarse-to-fine view of the entire video sequences...In this case, shots that are similar to each other are clustered together. Relations between clusters are governed by temporal ordering of shots within the two clusters..."; "From the clustering results and the temporal information associated with each shot, the system proceeds to build the graphs, with nodes representing scenes and edges representing the progress of the story from one scene to the next. The nodes capture the core contents of the video while the edges capture its structure" – col. 5, lines 36-41.

Therefore, Yeo et al. does disclose the clustering/classifying of shots in video sequences by using the shots' temporal information, and also the relations between clusters are governed by temporal ordering of shots within the clusters (cited cols. 4 and 5 above). Yeo also discloses "long sequences of related shots can be telescoped into a small number of **key frames which represent the repeatedly appearing shots** in the scene" – col. 2, lines 35-45. Because "key frames" represent "repeatedly appearing shots", a "key frame" does represent a class of shots. "Key frames" thus, are equivalent to the Appellant's limitation "key images". The limitation "temporal ordering of shots" seems to equivalent to "the ordering of shots' temporal values/distances". Thus, in the shots clustering process, similar shots and nodes will be clustered/merged together. This merging process in clustering/classifying objects is not novel in the technological art. Examiner combined Oguz's teaching with Yeo's teaching in order to show that using of temporal information/distance in clustering shots/frames in video sequences are not novel in the technological art. Oguz discloses in col. 8, lines 8-32 the detecting of scene change, the usage of TEMPORAL DISTANCE between the frames, the matching of edges between the frames etc…As Yeo discloses key frames represent the repeatedly appearing shots in the shots-clustering process – Yeo, col. 2, lines 35-45. Therefore, the usage of TEMPORAL DISTANCE to study scene change and to cluster shots are not novel in the technological art.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

Linh Black

/Linh Black/

Conferees:

Eddie Lee

James Trujillo

/James Trujillo/

Supervisory Patent Examiner, Art Unit 2169

/Eddie C. Lee/
Supervisory Patent Examiner, TC 2100